



In silico assessment of *EpCAM* transcriptional expression and determination of the prognostic biomarker for human lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC)

Abu Tayab Moin^a, Bishajit Sarkar^{b, **}, Md Asad Ullah^b, Yusha Araf^c, Nafisa Ahmed^d, Bashudev Rudra^{a, *}

^a Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chattogram, Bangladesh

^b Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh

^c Department of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh

^d Biotechnology Program, Department of Mathematics and Natural Sciences, School of Data and Sciences, BRAC University, Dhaka, Bangladesh

ARTICLE INFO

Keywords:

EpCAM gene
Lung adenocarcinoma
Lung squamous cell carcinoma
Gene expression
Biomarker
mRNA expression and regulation
Gene mutation

ABSTRACT

Epithelial cell adhesion molecule (*EpCAM*) is a transmembrane glycoprotein which is involved in cell signaling, proliferation, maturation, and movement, all of which are crucial for the proper development of cells and tissues. Cleavage of the *EpCAM* protein leads to the up-regulation of c-myc, e-fabp, and cyclins A and E which promote tumorigenesis. *EpCAM* can act as potential diagnostic and prognostic biomarker for different types of cancers as it is also found to be expressed in epithelia and epithelial-derived neoplasms. Hence, we aimed to analyze the *EpCAM* gene expression and any associated feedback in the patients of two major types of lung cancer (LC) i.e., lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), based on the publicly available online databases. In this study, server-based gene expression analysis represents the up-regulation of *EpCAM* in both LUAD and LUSC subtypes as compared to the corresponding normal tissues. Besides, the histological sections revealed the over-expression of *EpCAM* protein in cancerous tissues by depicting strong staining signals. Furthermore, mutation analysis suggested missense as the predominant type of mutation both in LUAD and LUSC in the *EpCAM* gene. A significant correlation (P-value < 0.05) between the higher *EpCAM* expression and lower patient survival was also found in this study. Finally, the co-expressed genes were identified with their ontological features and signaling pathways associated in LC development. The overall study suggests *EpCAM* to be a significant biomarker for human LC prognosis.

1. Introduction

The malady that ensues when cells adopt bizarre behaviors i.e., disregarding anti-growth signals, being indifferent to anti-apoptotic pathways, deceiving immune surveillance, and sustaining autonomous replicative potential and angiogenesis-is called cancer [1]. Such atypical cell growth is invasive and comes right after heart diseases as the leading cause of deaths worldwide. The International Agency for Research on Cancer (IARC) has reported a rise to 19.3 million cases and 10 million deaths due to cancer in 2020 [2]. Lung cancer (LC) is one of the most prominent cancers, comprising of 11.6% of the total cancer cases around the world, affecting both male and female alike [3]. According to the

Global Burden of Disease Study, the five-year survival rate of LC was 17.8% in 2020 which is significantly lower than that of other types of cancers [4]. And, LCs cause 18.4% of the total cancer deaths globally [5]. Non-small cell lung cancer (NSCLC) is the most common pathological type of LC that covers about 85% of all LCs. Squamous cell lung carcinoma (LUSC) and Lung adenocarcinoma (LUAD) are two important histopathological subtypes of NSCLC [6]. LUSC is one of the most common form of NSCLC where approximately 400,000 new cases occur annually, accounting for 20–30% NSCLCs. On the other hand, LUAD occurs more in non-smoking females and the survival rate of LUAD patients is poor because of drug resistance, failure to diagnose at an early stage and lack of efficient treatments [7–9]. Early diagnosis, especially

* Corresponding author.

** Corresponding author.

E-mail addresses: sarkarbishajit@gmail.com (B. Sarkar), bashudev@cu.ac.bd (B. Rudra).

<https://doi.org/10.1016/j.bbrep.2021.101074>

Received 12 May 2021; Received in revised form 2 July 2021; Accepted 5 July 2021

2405-5808/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

in the case of LUADs, can definitely reduce the mortality rate associated with LCs.

Previous studies suggest that the cell adhesion molecules like EpCAM protein have a role in cancer's intercellular and cell-extracellular matrix interactions. The differential expression i. e., upregulation or downregulation of several cell adhesion genes reveals the association between cell adhesion proteins and Differentially Expressed Genes (DEGs) [10]. For instance, *CDH-1* (Cadherin-1), *ITGB6* (Integrin beta 6) and *DSC3* (Desmocollin-3) are found to be upregulated in breast cancer [11]. Also, *CD44* and *CD29* (Integrin beta-1) are downregulated in ovarian cancer whereas, *ICAM-1* (Intercellular Adhesion Molecule 1) and *CDH3* (Cadherin-3) are upregulated [12]. Further, colorectal cancer upregulates *CDH1* and *CDH3*, but downregulates *CDH19* (Cadherin 19) and *PTPRF* [13]. LC has been shown to upregulate *EpCAM*, *ICAM-1*, *ITGA3* (Integrin alpha-3/beta-1), *ITGB4* (Integrin alpha-6/beta-4), *DSP* (Desmoplakin), and *DSC3* genes [14–16]. Thus, differential expression of DEGs could be used as possible biomarkers for detecting LCs like LUAD and LUSC [17].

A subset of cell adhesion proteins that are located on the surface of the cell and participate in cellular binding with either other cells or extracellular matrix (ECM) are called 'cell adhesion molecules (CAMs)' [18]. If a cell binds to another cell of the same type through these adhesion molecules, then this binding is called homotypic cell-cell adhesion [19,20]. Epithelial Cell Adhesion Molecule (EpCAM) is a trans-membrane glycoprotein of about 40 KD which consists of 314 amino acids. It is a homophilic cell-cell adhesion molecule independent of Ca²⁺ [21] and contains an extracellular domain (EpEX), a trans-membrane domain, and an intracellular domain (EpICD) [22]. According to both in vivo and in vitro studies, EpCAM protein has pivotal roles in cell signalling, proliferation, formation, differentiation, and maintenance of organ morphology [23]. EpCAM acts as a cell surface marker on progenitor cells as well as various stem cells along with being expressed in multiple types of epithelial tissues [24,25]. During the early stages of first development such as in the morula, the *EpCAM* gene is expressed, although it remains tissue-specific [26]. The murine homologue of *EpCAM* gene shows expression on thymocytes, antigen-presenting cells, T-cells, and definitely on epithelia. The interaction between intra-epithelial lymphocytes and epithelial cells is assisted by the selective expression of the CAMs on them [27]. Therefore, the expression of murine *EpCAM* of non-epithelial nature may facilitate homotypic adhesive interactions between dendritic cells or epithelial cells with thymocytes. Inconsistent expression like up-regulation or down-regulation or de novo expression is demonstrated by maximum CAMs during and after a malignant transformation occurs [26]. *EpCAM* is overexpressed in different types of cancers including breast cancer, ovarian cancer, and head and neck squamous cell cancer [28]. Due to the overexpression of EpCAM protein molecules in cancerous cells, it appears to be important in tumorigenesis and metastasis of carcinomas, so it can act as a potential prognostic marker as well as a possible target for immunotherapeutic strategies [29]. Previous studies revealed that *EpCAM* acts as prognostic marker in gallbladder carcinoma, breast cancer, and colorectal cancer. Moreover, *EpCAM* is also referred as a universal molecular marker for the detection of circulating tumor cells (CTCs) [30]. The overexpression of *EpCAM* has been proved to be an indicator of lesser survival rates in breast cancer and ovarian cancer patients [31,32]. This overexpression of *EpCAM* can be utilized as a biomarker for detection of LUAD and LUSC as *EpCAM* distributes differently for squamous cell carcinomas and adenocarcinomas. This adds a good diagnostic value in the precise recognition of a certain type of carcinoma [33,34].

LUAD and LUSC are the most prevalent of NSCLCs and the disease diagnosis in the early stages is unsatisfactory. Prognostic factors are meant to determine the features of patients and the stage of the cancer before the treatment is started. Some traditional prognostic factors for survival in patients with LUAD and LUSC are performance status, stage-tumor dimension, nodal status, and weight loss [35]. Moreover, symptoms are seen in advanced stages (stage III or stage IV) in about 70% of

patients diagnosed with LC [36]. The lack of symptoms in the initial stage and traditional prognosis such as determination of cancer metastasis by chest radiograph, CT scan or determining the presence of Napsin-A or TTF-1 proteins helps confirm the cancer progression. However, these methods are not satisfactory for early-prognosis and cannot lead to targeted treatment which is why survival rate is quite low [37]. Presently, there are almost no biomarkers available for detection of LC which can be applied in clinical use. This occurs either due to the lack of a robust sensitivity and specificity of the biomarkers or their functional relevance with lung carcinogenesis [36]. Thus, a biomarker is crucially required to detect LC in earlier stages, which can significantly increase the survival rate of the patients. Therefore, in this study, we hypothesize that studying gene expression of *EpCAM* may unveil its immunotherapy as well as prognostic value for diagnosis and treating LUAD and LUSC.

2. Materials and methods

2.1. Expression analysis of *EpCAM* in different types of cancerous and normal tissues

The expression pattern of *EpCAM* mRNA or transcript in numerous cancerous and their respective normal tissues was analyzed by three different databases i.e., Oncomine (<https://www.oncomine.org>) [38], GEPIA2 (<http://gepia2.cancer-pku.cn>) [39], and GENT2 (<http://gent2.appex.kr>) [40]. The Oncomine database is mainly used for analyzing the translational bioinformatics, which aids in discovering the highly ranked over-expressed genes from their 715 independent datasets, containing 86,733 clinical cancer samples and 12,764 standard tissue samples [38]. During the analysis, the threshold of p-value was kept at 1E-4 and the thresholds for fold change and gene rank were set at 2.0 and Top 10%, respectively. The Gene Expression Profiling Interactive Analysis or GEPIA2 webserver integrates the mRNA expression data of 9736 tumors and 8587 normal samples from The Cancer Genome Atlas (TCGA) program and the Genotype-Tissue Expression (GTEx) project. This server is widely recognized as the tool for differential expression analysis, patient survival analysis, detection of similar gene(s), correlation analysis etc. [39]. During analysis by GEPIA2 server, all the parameters were kept at their default values. And the Gene Expression database of Normal and Tumor tissues-2 (GENT2) server generates the result of expression analysis of a particular gene or mRNA by analyzing more than 68,000 samples stored in its database. This server has a user-friendly interface that uses the Apache Lucene indexing and Google Web Toolkit (GWT) framework to predict the results [40]. Like GEPIA2, during the experiment, all the parameters were kept default in GENT2 webserver.

2.2. Expression analysis of *EpCAM* in cancerous and normal lung tissues

Four different web servers were used to analyze the expression of *EpCAM* in normal and cancerous human lung tissues i.e., Oncomine [38], GEPIA2 [39], UALCAN (<http://ualcan.path.uab.edu/>) [41], and HPA (<https://www.proteinatlas.org/>) [42]. UALCAN is a user-friendly and interactive web server that allows the users to identify novel biomarkers and perform various in silico analyses of potential genes of interest by providing free access to the publicly available cancer OMICS data, such as the TCGA database [41]. This server was used to determine the relative expression pattern of *EpCAM* in both the target LUAD and LUSC samples from the TCGA database. The Human Protein Atlas (HPA) is a publicly available, Swedish-based project which was initiated in 2003. The aim of this project is to map all the human proteins in cells, tissues and organs using an integration of different types of omics technologies, for example, mass spectrometry-based proteomics, transcriptomics, antibody-based imaging etc. [42]. The HPA database is the official webserver of this project. In the HPA database, a visual comparison was made between a cancerous lung tissue (with lung-specific

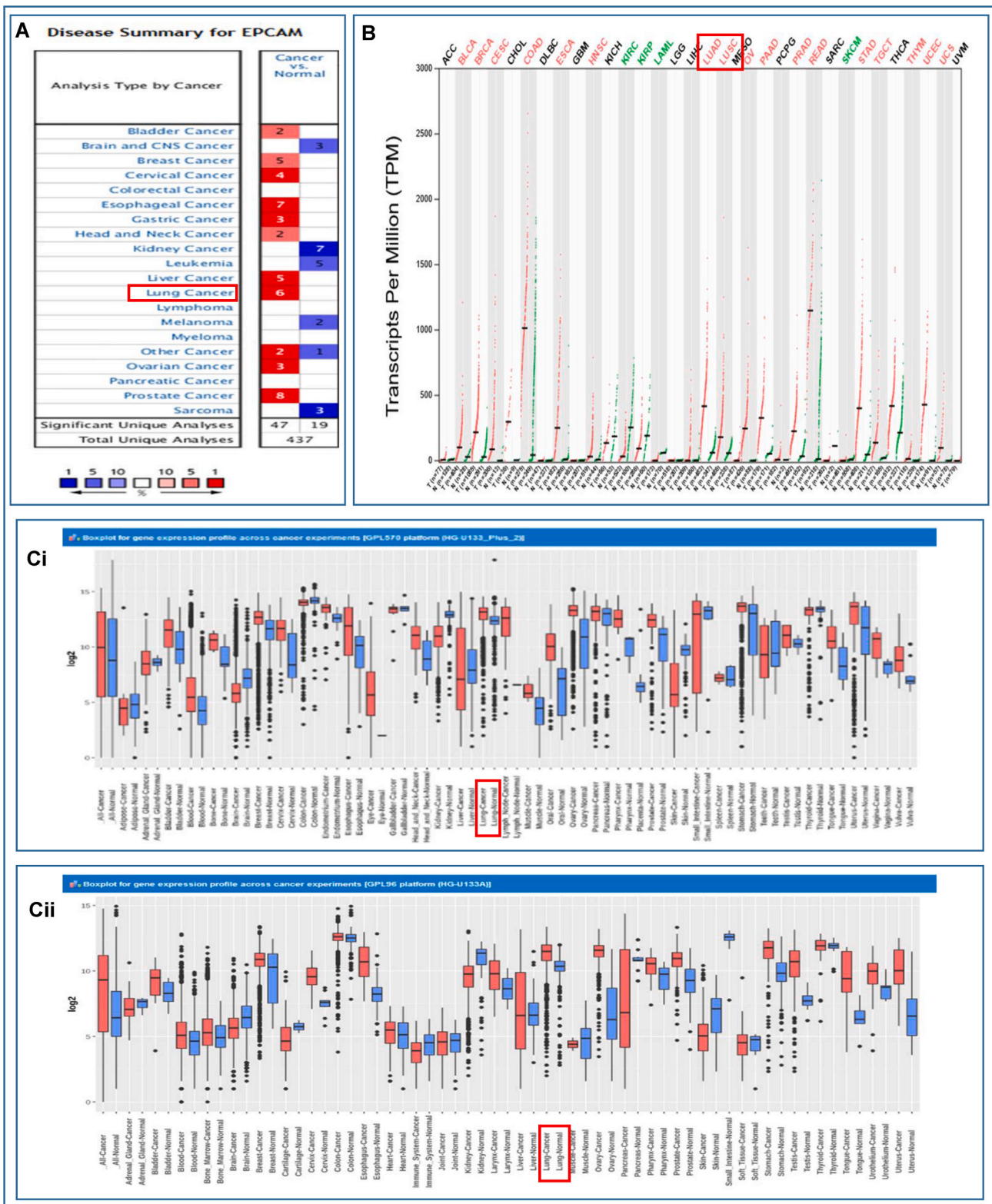


Fig. 1. Pattern of the tissue-wide expression of *EpCAM* in different human cancers, (A) comparison between cancer versus normal tissues in which high and low expression of mRNA has been indicated by red and blue colors, respectively, (B) the dot plot depicts the gene expression profile of the *EpCAM* gene in 33 different types of human cancers including tumors and normal tissue samples together. Herein, red and green dashed lines indicate the average expression value in all tumor and normal tissues, respectively, and (C) the box-plot showing the *EpCAM* mRNA expression in tumors and respective normal tissues using the Affymetrix HG-U133pLUS2 (Ci) and HG-U133A platforms (Cii) of the GENT2 database where the boxes indicate the median, the dots represent outliers, the red-boxes refer to the tumor tissues, whereas the blue-boxes represent the expression of normal tissues. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

antibody HPA026761) and a normal lung tissue (without any lung-specific antibody CAB030012). In all these servers, default parameters were used to predict the results and the p-value less than 0.05 was considered statistically significant.

2.3. Analysis of the association of *EpCAM* expression with different clinical features and promoter methylation

To analyze the relationship of *EpCAM* gene expression with different types of clinical features, the UALCAN server [41] was again used in this step. In this study, both the LUAD and LUSC cancerous tissues were compared with normal tissues and numerous features i.e., individual cancer stages, patient's race, patient's gender, patient's age, patient's smoking habit, tumor histology, nodal metastasis status, and T53 mutation status, were noted. Thereafter, to analyze the DNA methylation pattern, the UCSC Xena Functional Genomic Explorer (<https://xenabrowser.net/>) was utilized which provides the users to analyze the functional genomic data for determining the correlations between numerous genomic and phenotypic variables [43]. The server generated the methylation pattern of *EpCAM* gene by analyzing 877 LUAD and 765 LUSC samples from GDC TCGA database. From the server, the Illumina Human Methylation 27 and Illumina Human Methylation 450 patterns were analyzed keeping all the parameters default.

2.4. Analysis of mutations and copy number alterations in the *EpCAM* gene

The cBioportal server (<https://www.cbioportal.org/>) was used to determine the pattern of the mutations and alterations in copy number of the *EpCAM* gene. The server is a freely available online server designed to visualize and interpret multidimensional cancer genomics datasets [44]. In our analysis, total 2983 LUAD samples and 1176 LUSC samples from different databases available in the server, were utilized to find out the anomalies or mutations and copy numbers of the *EpCAM* gene using the GISTIC algorithm (Genomic Identification of Significant Targets in Cancer), keeping all the parameters default. The server predicts the probable mutations in different regions along the *EpCAM* gene, which might be responsible for cancer development.

2.5. Determining the relationship between *EpCAM* gene expression and the survival of LC patients

The relationship between the expression of *EpCAM* gene and the survival of patients with LC was determined by the Prognoscan server (<http://dna00.bio.kyutech.ac.jp/Prognoscan/>). Prognoscan server is a database for meta-analysis of the prognostic values of genes. It searches the publicly available microarray datasets to establish a relationship between the expression of a target gene and the prognosis of the disease caused by that particular gene such as overall survival (OS) and disease free survival (DFS) [45]. The server plots the expression levels of a gene using the Kaplan-Meier statistical method. During the analysis, default parameters were used and any p-value <0.05 was considered as the statistically significant value.

2.6. Determination of the genes co-expressed with *EpCAM* in LC tissues

The profiling of genes co-expressed with *EpCAM* was performed by three different servers i.e., OncoPrint [38], GEPIA2 [39], and UCSC Xena web browser [43]. At first, the OncoPrint server was used to find out the co-expressed genes. The server generates the correlation scores and the co-expressed genes are expected to have better correlation scores. Thereafter, the GEPIA2 server was exploited to analyze the correlation between *EpCAM* gene and the gene that was found to be highly co-expressed with *EpCAM* in the OncoPrint server. GEPIA2 server generated the p-values when presenting the relation of the two selected genes and any p-value with less than 0.05 value was considered as

Table 1

The expression of *EpCAM* in the two subtypes of LC from the OncoPrint database.

Dataset	Parameters	Samples	P-value	Gene Rank	Fold Change
Talbot Lung (n = 93)	Normal Lung	2	–	–	–
	Squamous Cell Lung Carcinoma	34	1.94E-9	302	4.253
	Adenocarcinoma	30	–	–	–
Su Lung (n = 66)	Normal Lung	27	1.29E-6	292	2.157
	Adenocarcinoma	19	–	–	–
	Normal Lung	20	1.27E-8	72	2.337
Strearman Lung (n = 39)	Adenocarcinoma	58	–	–	–
	Normal Lung	58	2.18E-20	124	2.413
	Adenocarcinoma	49	–	–	–
Landi Lung (n = 107)	Normal Lung	58	3.09E-19	100	2.197
	Adenocarcinoma	10	–	–	–
	Normal Lung	86	1.46E-10	27	2.569
Beer Lung (n = 96)	Adenocarcinoma	10	–	–	–

statistically significant. Finally, the UCSC Xena web browser was used to predict the gene expression pattern of the two selected genes in LC patients. The co-expression analysis was conducted for both LUAD and LUSC.

2.7. Determination of gene ontology and signaling pathways of *EpCAM* and its related genes in LC development

The gene ontology (GO) and cell signaling pathways of *EpCAM* gene were retrieved from the Enrichr server (<http://amp.pharm.mssm.edu/Enrichr/>). This server takes the advantage of using the enrichment analysis inferring knowledge about a target gene set by comparing it to multiple genomics datasets representing prior biological knowledge [46, 47]. In this analysis, we used the *EpCAM* gene and all the co-expressed genes from the previous step to find out the Go terms i.e., GO biological process, GO molecular function, and GO cellular component. And the signaling pathways were also determined from BioPlanet 2019, KEGG 2019 Human, and Reactome 2016 databases via the Enrichr web server.

3. Results

3.1. Analysis of the mRNA expression of *EpCAM* in different types of cancerous and normal tissues

The differential expression of *EpCAM* transcripts or mRNAs in different cancer tissues and their respective healthy counterparts was determined using three web-based servers, i.e., OncoPrint, GEPIA2, and GENT2 servers. In the OncoPrint server, the statistically significant results were found in only 66 studies ranked in the top 10%, among the 437 unique datasets. The up-regulation of *EpCAM* was evident in numerous malignancies including lung, breast, bladder, cervical, esophageal, gastric, head and neck, liver, ovarian, prostate, and other cancers. Likewise, *EpCAM* was found to be highly expressed in both LUAD and LUSC patients while downregulated in the patients of Brain, CNS, Kidney Cancer, Leukemia, Melanoma, and Sarcoma (Fig. 1A). Fig. 1B illustrates the comparative expression of *EpCAM* in 33 different human cancers and their corresponding normal tissues. Furthermore, the overexpression of *EpCAM* in various cancers was revealed in another analysis using the Affymetrix HG-U133PLUS2 (Fig. 1Ci) and HG-U133A (Fig. 1Cii) platform of the GENT2 database as portrayed in Fig. 1C. Therefore, these outcomes provided strong shreds of evidence regarding the higher expression of *EpCAM* gene in the LC tissues when contrasted with the normal, non-cancerous tissues (Fig. 1).

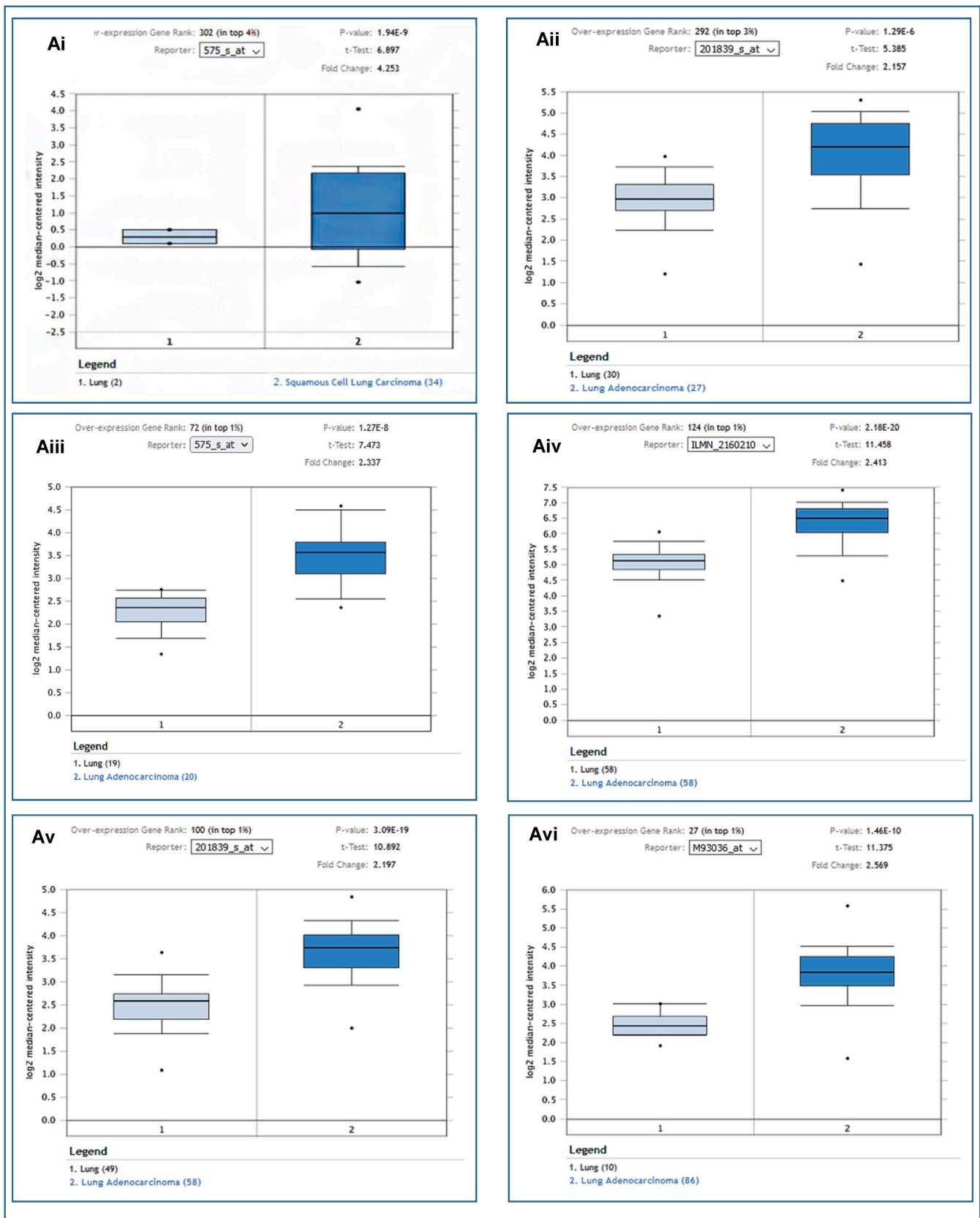


Fig. 2. Analysis of the *EpCAM* expression in LUAD and LUSC in which (A) box-plots showing comparative expression between normal (left) and cancer tissue (right) - for LUAD and LUSC (Ai-Avi), (B-C) box-plots showing the expression of *EpCAM* mRNA based on sample types in tumor tissue and normal tissues for LUAD (Bi) and LUSC (Bii), using the UALCAN and GEPIA2 (C) servers, respectively, and (D-F) the immunohistochemistry images of the expression of *EpCAM* in LUAD (D) and LUSC (E) tissues as well as normal tissue (F) retrieved from the HPA database. The number within the parenthesis in Fig. 2 (Ai-Avi) represents the number of the samples deposited in the database.

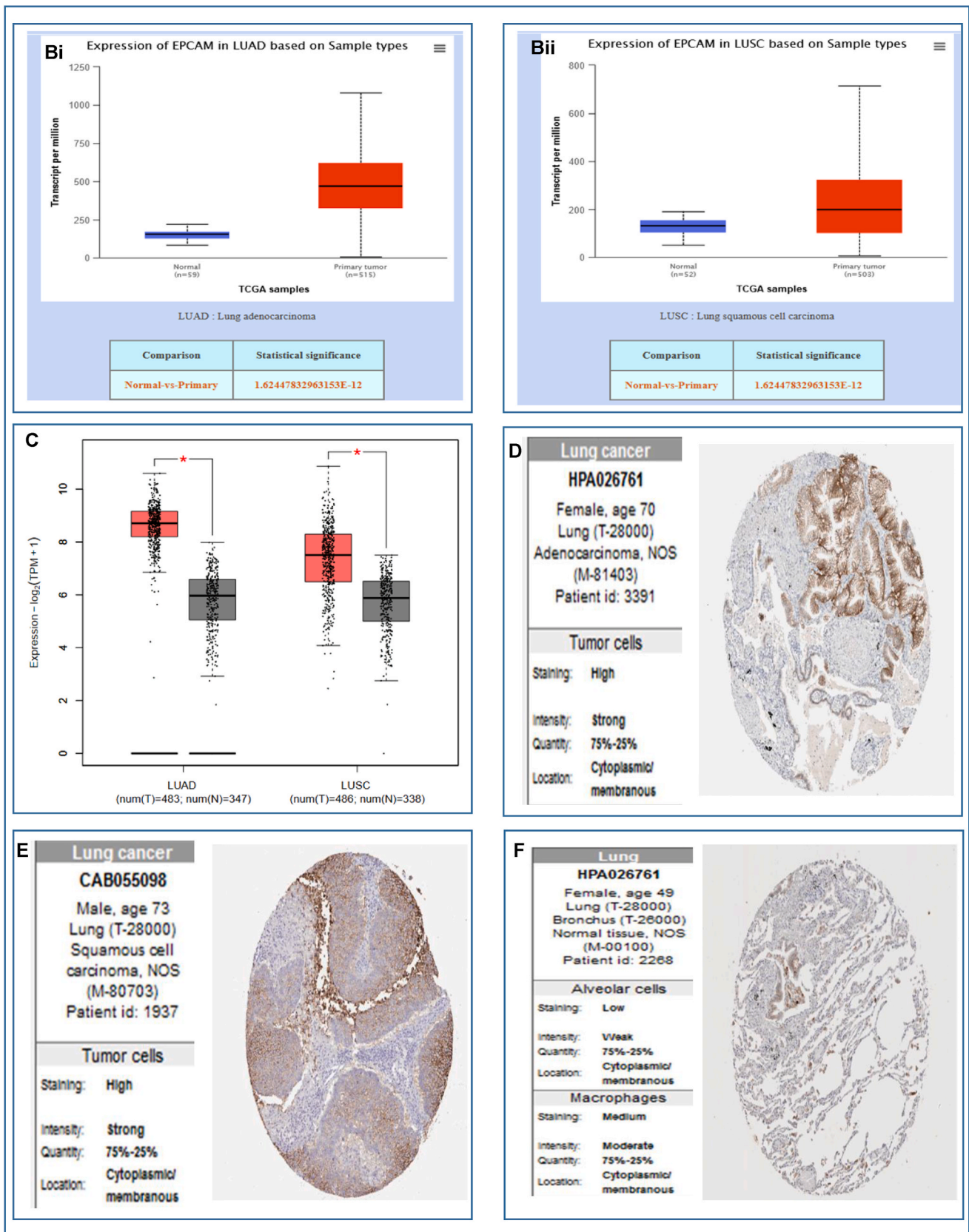


Fig. 2. (continued).

3.2. Expression of *EpCAM* transcript in human LC tissues

The OncoPrint webserver was used to analyze the *EpCAM* gene

expression for each subtype of LC compared to the normal tissues. The results revealed the overexpression of *EpCAM* in two different LCs i.e., LUAD and LUSC (Fig. 2Ai-ii and Table 1). Further assessment of TCGA

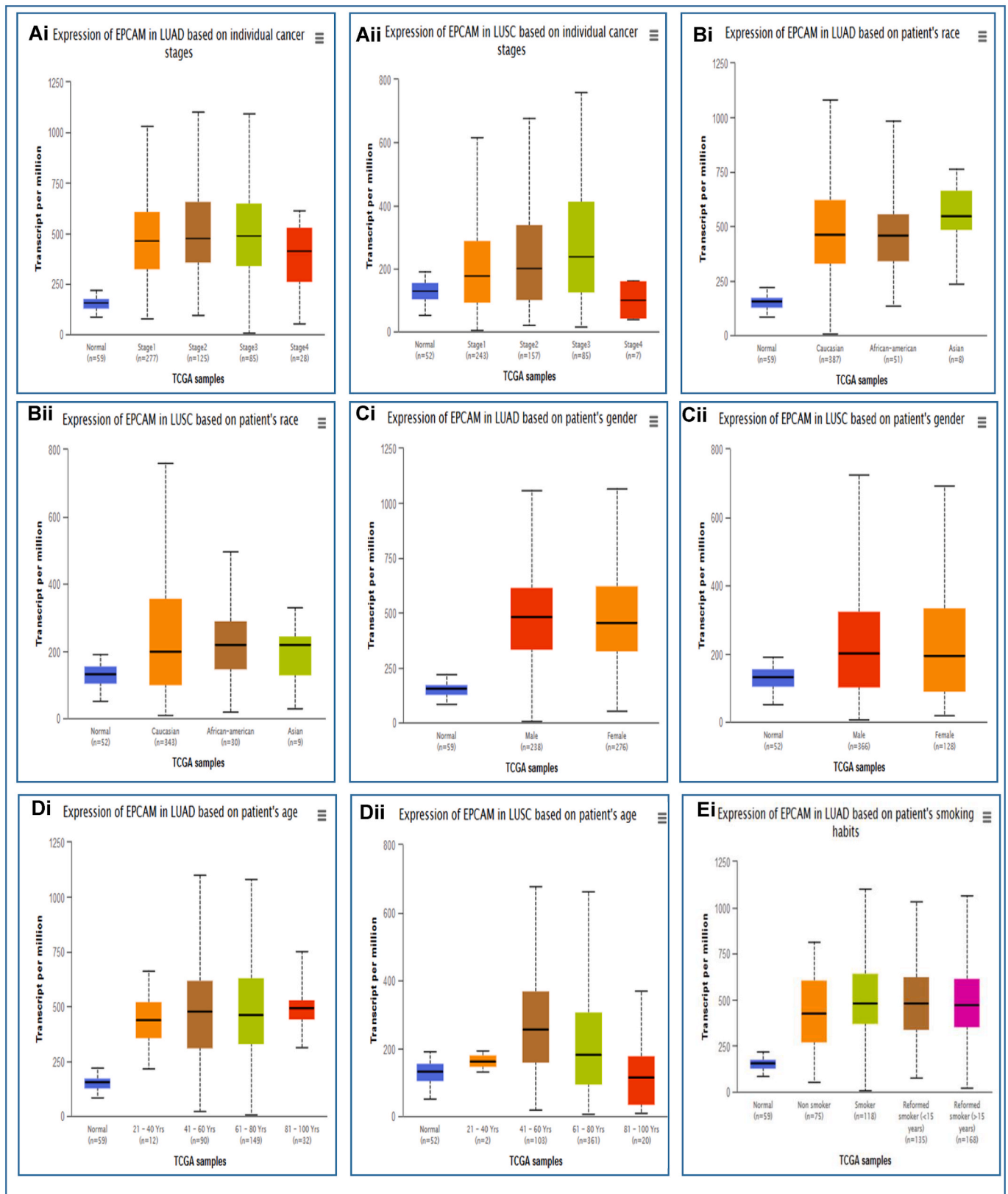


Fig. 3. The analysis of *EpCAM* expression with clinical characteristics of LUAD and LUSC patients. The *EpCAM* mRNA expression in LUAD (Ai-Hi) and LUSC (Aii-Hii) showing individual cancer stage, patient's race, gender, age group, smoking habit, histological subtypes, nodal metastasis, and TP53 status, respectively.

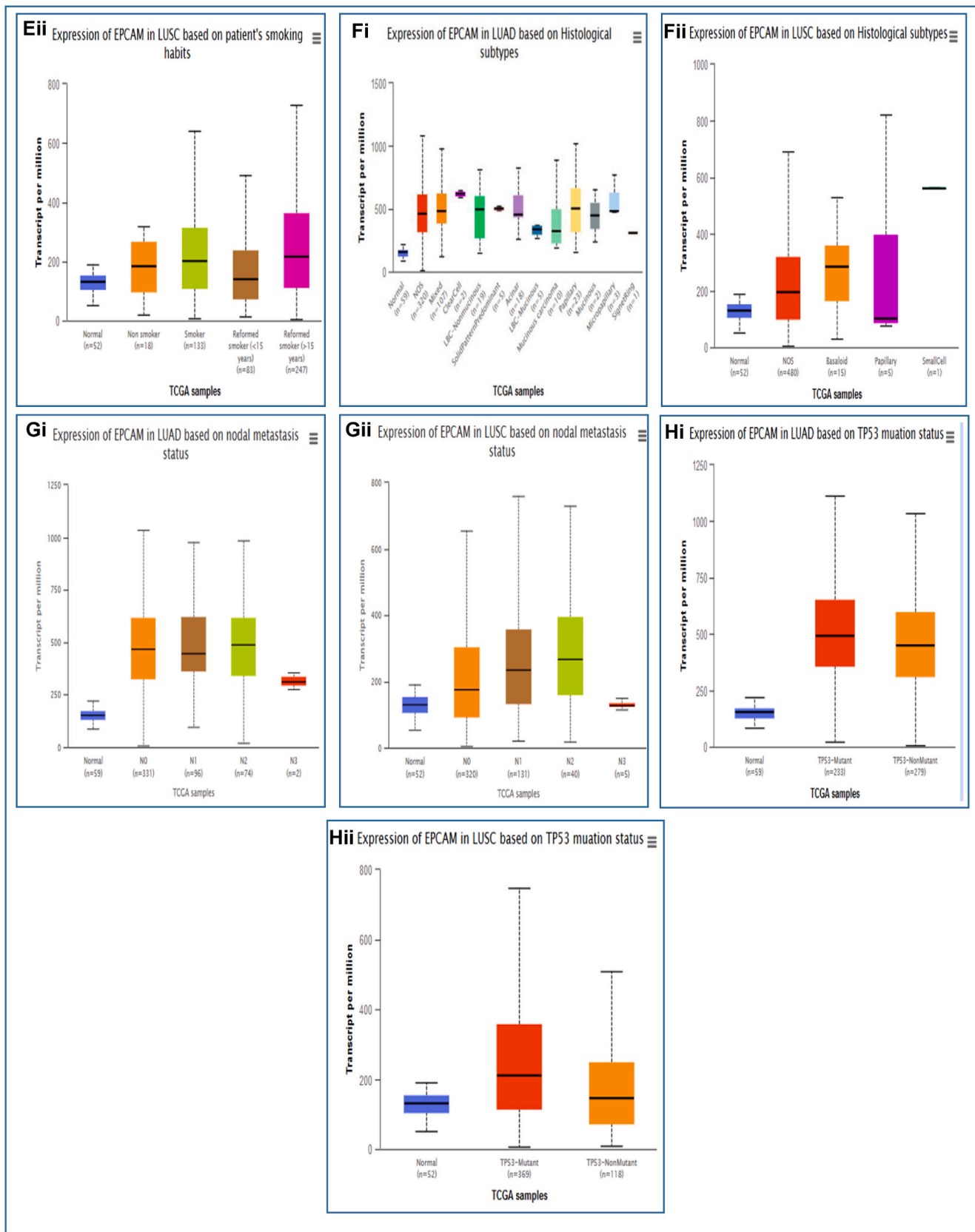


Fig. 3. (continued).

datasets with UALCAN and GEPIA2 servers also showed considerable up-regulation of *EpCAM* in both LUAD and LUSC (Fig. 2Bi, 2Bii and 2C). In addition to that, the relative immunohistochemistry analysis was

performed between normal and cancer tissues using the HPA database. While comparing the staining signals between normal and cancerous lung tissues, the signal was found to be moderate to weak in normal

Table 2

aThe relationship between *EpCAM* and numerous clinicopathological features of LUAD.

<i>EpCAM</i> expression based on	Features	Expression of mRNA	Number of samples	Statistical significance (p-value)
Individual cancer stages	Normal	Underexpression	59	–
	Stage-1	Overexpression	277	<1E-12
	Stage-2	Overexpression	125	1.624E-12
	Stage-3	Overexpression	85	<1E-12
Patient's race	Stage-4	Overexpression	28	2.237E-06
	Normal	Underexpression	59	–
	Caucasian	Overexpression	387	<1E-12
	African American	Overexpression	51	1.728E-12
Patient's gender	Asian	Overexpression	8	2.996E-04
	Normal	Underexpression	59	–
	Male	Overexpression	238	1.624E-12
Patient's age	Female	Overexpression	276	<1E-12
	Normal	Underexpression	59	–
	21–40 Yrs	Overexpression	12	5.362E-02
	41–60	Overexpression	90	1.624E-12
	61–80	Overexpression	149	<1E-12
Patient's smoking habit	81–100	Overexpression	32	2.480E-10
	Normal	Underexpression	59	–
	Non smoker	Overexpression	75	<1E-12
	Smoker	Overexpression	118	<1E-12
	Reformed	Overexpression	135	<1E-12
	Smoker 1	Overexpression	168	1.624E-12
Histological subtypes	Reformed Smoker 2	Overexpression	168	1.624E-12
	Normal	Underexpression	59	–
	NOS	Overexpression	820	1.624E-12
	Mixed	Overexpression	107	1.624E-12
	Clear Cell	Overexpression	2	<1E-12
	LBC-Non Mucinous	Overexpression	19	2.341E-06
	Solid Pattern Predominant	Overexpression	5	1.921E-03
	Acinar	Overexpression	18	5.900E-05
	LBC-Mucinous	Overexpression	5	4.740E-02
	Mucinous carcinoma	Overexpression	10	7.959E-03
	Papillary	Overexpression	23	8.074E-08
	Mucinous	Overexpression	2	3.905E-01
	Micropapillary	Overexpression	3	5.009E-02
	Nodal metastasis status	Normal	Underexpression	59
NO		Overexpression	331	1.624E-12
N1		Overexpression	96	1.624E-12
N2		Overexpression	74	<1E-12
N3		Overexpression	2	3.781E-08
TP53 mutation status	Normal	Underexpression	59	–
	TP53-Mutant	Overexpression	233	1.624E-12
	TP53-NonMutant	Overexpression	279	<1E-12

2b. The relationship between *EpCAM* and numerous clinicopathological features of LUSC.

<i>EpCAM</i> expression based on	Features	Expression of mRNA	Number of samples	Statistical significance (p-value)
Individual cancer stages	Normal	Underexpression	52	–
	Stage-1	Overexpression	243	1.3E-12
	Stage-2	Overexpression	157	3E-13
	Stage-3	Overexpression	85	1.232E-10
Patient's race	Stage 4	Overexpression	7	5.25E-01
	Normal	Underexpression	52	–
	Caucasian	Overexpression	343	1.11E-16
	African American	Overexpression	30	2.375E-04
Patient's gender	Asian	Overexpression	9	1.351E-01
	Normal	Underexpression	52	–
	Male	Overexpression	366	<1E-12
Patient's age	Female	Overexpression	128	1.118E-10
	Normal	Underexpression	52	–
	21–40 Yrs	Overexpression	2	1.683E-01
41–60	Overexpression	103	1.743E-12	

Table 2 (continued)

2b. The relationship between *EpCAM* and numerous clinicopathological features of LUSC.

<i>EpCAM</i> expression based on	Features	Expression of mRNA	Number of samples	Statistical significance (p-value)
Patient's smoking habit	61–80	Overexpression	361	1.624E-12
	81–100	Overexpression	20	4.949E-01
	Normal	Underexpression	52	–
	Non smoker	Overexpression	18	9.979E-03
	Smoker	Overexpression	133	1.36E-10
	Reformed Smoker 1	Overexpression	83	6.791E-04
Tumor histology	Reformed Smoker 2	Overexpression	247	1.11E-16
	Normal	Underexpression	52	–
	NOS	Overexpression	480	<1E-12
Nodal metastasis status	Basaloid	Overexpression	15	2.115E-03
	Papillary	Overexpression	5	3.078E-01
	Normal	Underexpression	52	–
	NO	Overexpression	320	1.624E-12
	N1	Overexpression	131	4.251E-13
TP53 mutation status	N2	Overexpression	40	2.341E-06
	N3	Overexpression	5	2.815E-01
	Normal	Underexpression	52	–
	TP53-Mutant	Overexpression	369	<1E-12
TP53 mutation status	Mutant	Overexpression	118	5.822E-05
	TP53-NonMutant	Overexpression	118	5.822E-05

alveolar cells, whereas both LUAD and LUSC tissue samples showed strong staining signals as depicted in Fig. 2D, E and 2F, respectively. The results indicate that the *EpCAM* expression is relatively higher in cancerous tissues than the normal tissues (Fig. 2 and Table 1).

3.3. Association of *EpCAM* expression with clinical characteristics of LC patients

The relationship between *EpCAM* gene expression and the clinical representation of LUAD and LUSC patients along with the control healthy individuals was observed using the TCGA dataset via the UALCAN database. Up-regulation of *EpCAM* for LUAD resulted in different parameters including individual cancer stages, patient's race, gender, age, smoking habit, tumor histology, nodal metastasis, and TP53 mutation status, as depicted in Fig. 3 and listed in Table 2a. In the analysis, the increased expression of *EpCAM* was the highest in a fluctuating manner for LUAD (Fig. 3Ai) and in an incremental manner for LUSC (Fig. 3Aii) in the middle stages of LC. Also, the expression of *EpCAM* was found to be the highest among the patients from Asia for LUAD (Fig. 3Bi) and Caucasian regions for LUSC (Fig. 3Bii). The increased expression level was almost the same in male and female patients for both LUAD and LUSC (Fig. 3Ci, ii). Moreover, Fig. 3Di and ii shows higher expression in the middle age group (41–80 years old) for LUAD, whereas the expression was higher in the age group of 41–60 years old for LUSC. Up-regulation of *EpCAM* expression can also be seen considering other clinicopathological parameters, including smoking habit (Fig. 3Ei, ii), tumor histology (Fig. 3Fi,ii), nodal metastasis (Fig. 3Gi,ii), and TP53 mutation status (Fig. 3Hi, ii) (Table 2ab). These results indicate that the expression of *EpCAM* and clinical characteristics is significantly higher in LC compared to the healthy, normal individuals.

3.4. Analysis of promoter methylation of LC from TCGA dataset

DNA methylation is crucial to understand because when it is found in a gene promoter, it acts to inhibit or lower the transcription process. Thus, promoter methylation is an important factor to be considered while studying cellular development, gene silencing, mRNA expression, tissue differentiation, genetic imprinting etc [48]. Also, hypermethylation of high-density CpG regions or genome-wide

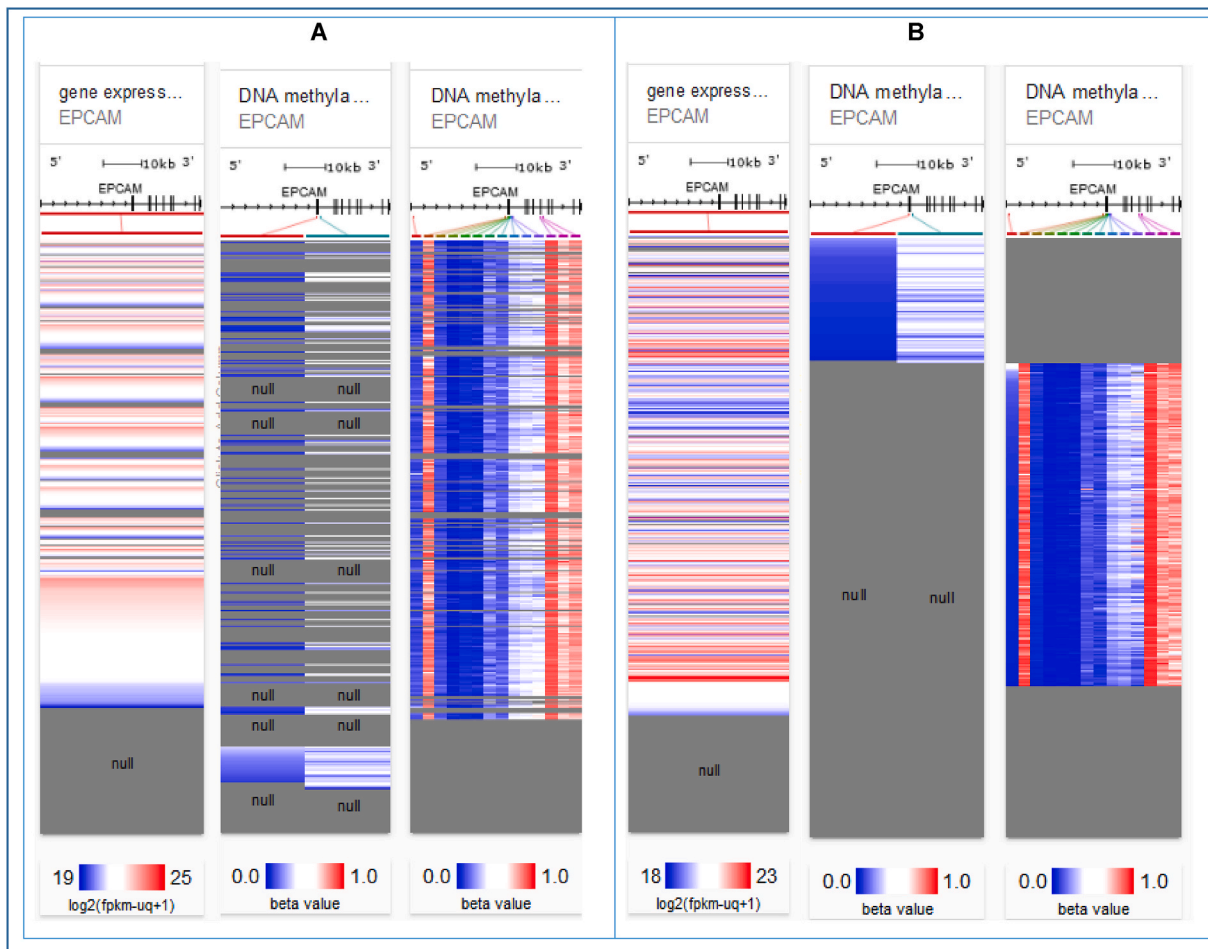


Fig. 4. Promoter methylation of the *EpCAM* gene in LC tissues. Heat map of *EpCAM* expression and DNA methylation status for LUAD (A) and LUSC (B).

Table 3

A list of *EpCAM* mutational positions and types in LUAD and LUSC from the TCGA dataset.

Sample ID	Cancer Type	Protein Change	Mutation Type	Chromosome No	Start Position	End Position	Number of Samples
TCGA-78-8662-01	Lung Adenocarcinoma	X62_splice	Splice	2	47600710	47600710	1014
LUAD-B00416	Lung Adenocarcinoma	G222V	Missense	2	47606915	47606915	803
LUAD-S01315	Lung Adenocarcinoma	N91S	Missense	2	47601034	47601034	1765
BGI-RS55	Lung Adenocarcinoma	Q204E	Missense	2	47606146	47606146	107
TCGA-44-4112-01	Lung Adenocarcinoma	G263V	Missense	2	47607038	47607038	371
TCGA-44-6779-01	Lung Adenocarcinoma	A210T	Missense	2	47606164	47606164	89
TCGA-78-7163-01	Lung Adenocarcinoma	D98E	Missense	2	47601056	47601056	24
TCGA-44-4112-01	Lung Adenocarcinoma	G263V	Missense	2	47607038	47607038	434
TCGA-44-6779-01	Lung Adenocarcinoma	A210T	Missense	2	47606164	47606164	91
TCGA-78-7163-01	Lung Adenocarcinoma	D98E	Missense	2	47601056	47601056	49
TCGA-86-8279-01	Lung Adenocarcinoma	E137A	Missense	2	47601172	47601172	458
TCGA-44-4112-01	Lung Adenocarcinoma	G263V	Missense	2	47607038	47607038	389
TCGA-44-6779-01	Lung Adenocarcinoma	A210T	Missense	2	47606164	47606164	93
TCGA-78-7163-01	Lung Adenocarcinoma	D98E	Missense	2	47601056	47601056	27
TCGA-37-3789-01	Lung Squamous Cell Carcinoma	M294I	Missense	2	47612328	47612328	379
TCGA-98-8021-01	Lung Squamous Cell Carcinoma	E147Q	Missense	2	47602386	47602386	930
TCGA-22-4604-01	Lung Squamous Cell Carcinoma	P3S	Missense	2	47596651	47596651	225
TCGA-37-3789-01	Lung Squamous Cell Carcinoma	M294I	Missense	2	47612328	47612328	977

hypomethylation has been proved to be correlated with numerous types of cancer. Hence, the correlation between *EpCAM* expression and DNA methylation for both LUAD and LUSC was analyzed using two different methylation assays available in the server i.e., Illumina Human Methylation 27 and Illumina Human Methylation 450. The result in the heat maps revealed a likewise negative relation between the *EpCAM* expression and some CpG islands which indicates the overexpression of *EpCAM* and for both genes methylation was found to be responsible for

mutation and possibly, cancer formation (Fig. 4).

3.5. Analysis of mutations, copy number alterations, and expression of mutant *EpCAM* transcript

Alterations of the *EpCAM* gene in both types of LCs covering a total of 2983 samples from 8 LUAD studies, and 1176 samples from 3 LUSC studies were determined using cBioPortal database (Table 3). The

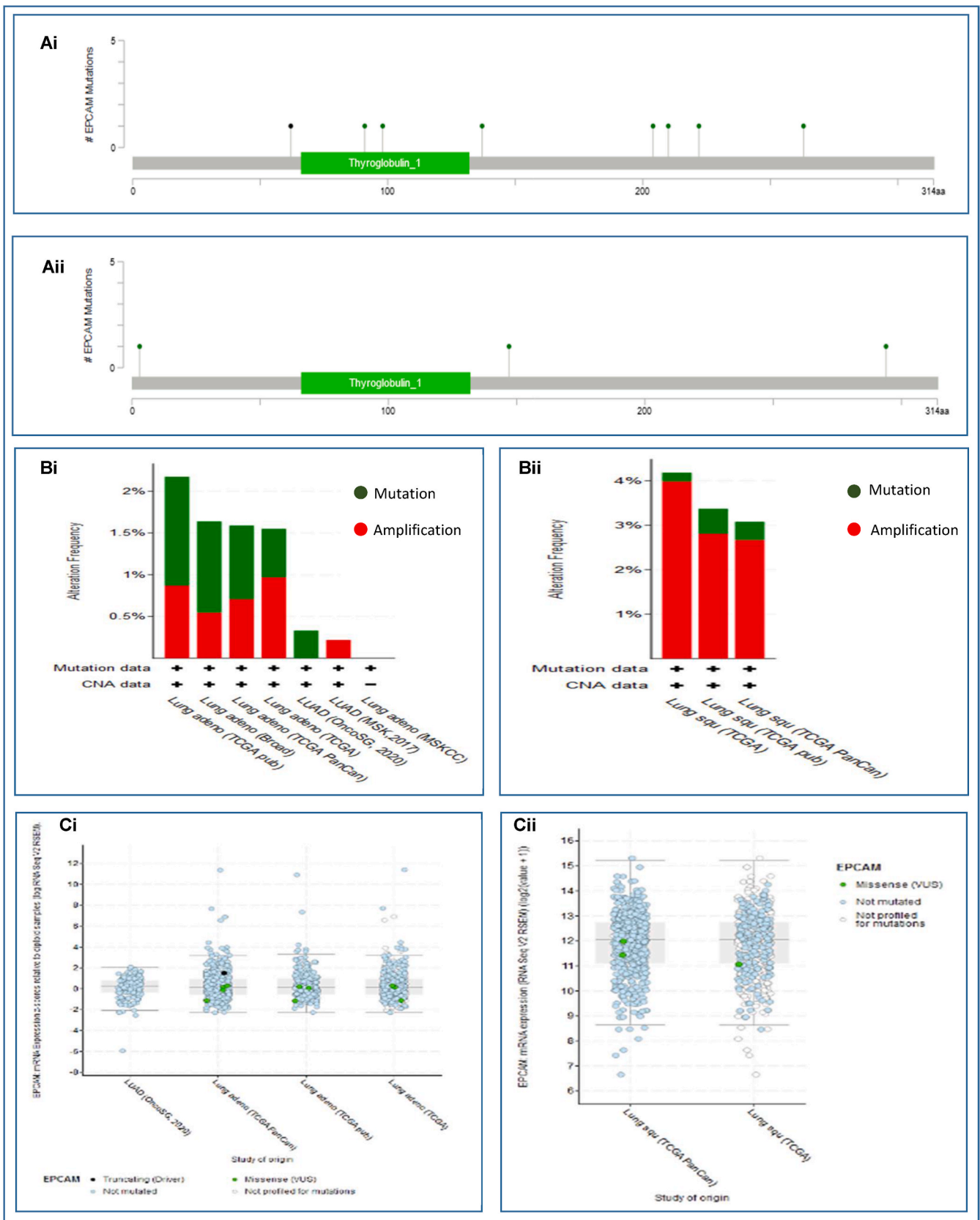


Fig. 5. Genetic alteration and mutations of *EpCAM* in LUAD and LUSC tissues. Herein (Ai) lollipop plot shows the type of alteration in fourteen mutation spots within the peptide sequence (1–314 residues) of Thyroglobulin_1 domain in *EpCAM* from LUAD tissues, (Aii) depicts the alterations in only four mutation spots in *EpCAM* from LUSC tissues, (Bi and Bii) bar diagrams show the mutation frequencies and genome alteration in the *EpCAM* gene for LUAD and LUSC, respectively, and (Ci and Cii) indicate the correlation between the expression and copy number alteration of *EpCAM* for LUAD and LUSC in the TCGA dataset.

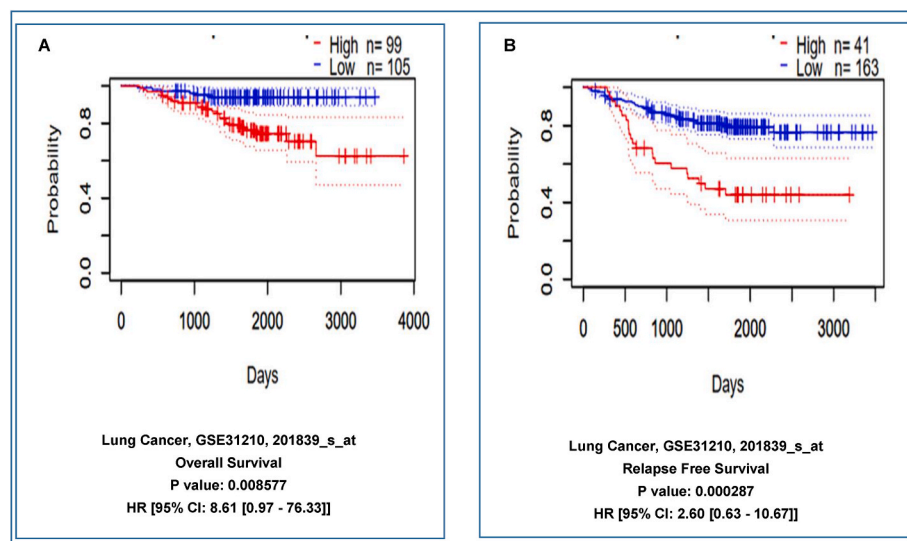


Fig. 6. The in Kaplan-Meier plot represents the relationship between *EpCAM* gene expression and survival of LC patients. The survival curves demonstrate patients' survival with the high (red) and low (blue) expression of *EpCAM* in the plots where (A–B) showing overall survival, and relapse-free survival, respectively. The analysis was focused on the *EpCAM* expression in LC patients. Here, HR, hazard ratio; CI, confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

EpCAM gene was found to be altered in 28 (<1%) among the quarried LUAD samples with an achieved somatic mutation frequency of 0.5%. A total of 14 mutations including 13 missenses and 1 splice mutations located within 1–314 residues of Thyroglobulin_1 domain were identified. All these mutations took place within 47600710 to 47607038 base pair of chromosome 2. Interestingly, the highest mutation in both LUAD and LUSC was the missense mutation, occurring in a total of 1765 (Fig. 5Ai) and 977 samples (Fig. 5Aii), respectively. Furthermore, in the alteration frequency analysis, LUAD TCGA pub occupied the highest frequency of 2.17% of 230 cases, among the six categories of the studies (Fig. 5Bi). Finally, profiling of mutated *EpCAM* mRNA expression was conducted where eleven cases of mutation including ten missense and one non-sense were found in LUAD (Fig. 5Ci). Hence, the results suggested that alterations found in *EpCAM* protein in the quarried LUAD tissues are not possibly responsible for the over expression of *EpCAM*. Again, the *EpCAM* gene was altered in 42 (4%) of quarried LUSC samples with a somatic mutation frequency of 0.3%. A total of 4 missense mutations were detected in patients with multiple samples of LUAD. In the alteration frequency analysis, the highest frequency was reported (3.37% of 178 cases) for LUSC TCGA pub among the three categories of the studies (Fig. 5Bii). Thereafter, the mutated *EpCAM* mRNA expression was profiled for LUSC and the result revealed three cases of analysis all of which were missense mutations (Fig. 5Cii). Therefore, these findings suggested that the overexpression of *EpCAM* in LUSC might not be correlated with mutations or copy number alterations found in *EpCAM* protein just like the LUAD samples (Fig. 5). This also provided indication that mutations or alterations in *EpCAM* gene might not be pivotal in causing LC even though this could be a case that is needed further studies.

3.6. *EpCAM* expression and clinical prognosis of LC patients

The relationship between the level of *EpCAM* expression and patient's survival in LC was analyzed using the Prognoscan database (significant level was kept at P-value < 0.05 and HR > 1). The analysis showed a negative correlation of *EpCAM* expression with the survival of the patients. According to the analysis, high expression of *EpCAM* might lead to lower survival rate whereas low or normal expression of the protein should enhance the survival rate of the patients. For dataset GSE31210 (Number of samples = 204), patients with low *EpCAM* expression (n = 105 and 163) had significantly higher survival probability whereas higher expression (n = 99 and 41) of *EpCAM* was responsible lower survival rate in the patients as illustrated in Fig. 6A and B. The results revealed that the enhanced expression of *EpCAM*

could be responsible for relatively poor prognosis in LC patients.

3.7. Analysis of gene signatures linked to *EpCAM* and human LC

The co-expression profile of *EpCAM* was analyzed with 19 genes using a total of 16 samples from LUAD and LUSC patients through the Oncomine database (Fig. 7A). The *HOXB7* was found to be mostly co-expressed (R = 0.998) among the total 19 genes. The *EpCAM* and *HOXB7* were found to be positively correlated with the Spearman coefficient in both LUAD (R = 0.14) (Fig. 7Bi) and LUSC (R = 0.26) (Fig. 7Bii) using GEPIA2 server. Furthermore, the Pearson and Spearman correlation analyses were performed to ensure the positive correlation between *EpCAM* and *HOXB7* in LC patients using TCGA data through UCSC Xena server (Fig. 7C–D). For the LUAD (Fig. 7Ci–Di) and LUSC (Fig. 7Cii–Dii) patients, the Pearson correlation values were found to be 0.1243 and 0.2117, respectively. Besides this, the Spearman correlation values were also reported to be 0.1513 and 0.2894, respectively. Therefore, these results suggest that these two genes might be inter-linked in different signaling pathways in LC progression.

3.8. Determination of gene ontologies and signaling pathways linked to *EpCAM* and LC progression

Based on the *EpCAM* and correlated genes, signaling pathways and gene ontological features that lead to the progression of LC in humans were identified. For pathway determination, the results from three databases depicted in Fig. 8A–C were considered. In the analysis of KEGG human 2019 database, different significant pathways were found including mucin type O-glycan biosynthesis and branched chain amino acid (BCAA) metabolism i.e., degradation of valine, leucine and isoleucine. (Fig. 8A). Furthermore, Reactome 2016 database showed pathways related to amino acid catabolism, protein metabolism, activation of gene expression by RORA, BMAL1:CLOCK, NPAS2, and SREBF (SREBP), YAP1- and WWTR1 (TAZ)- stimulated gene expression, mitochondrial biogenesis, regulation of transcription by NOTCH1 intracellular domain, etc. (Fig. 8B). Finally, the analysis of the Bioplanet 2019 database revealed the p38 alpha/beta MAPK downstream pathway as well as some other significant pathways related to aflatoxin B1 metabolism, benzo(a)pyrene metabolism, amino acid catabolism, Rho-mediated activation of SRF, eicosanoid metabolism, YAP1- and WWTR1 (TAZ)-stimulated gene expression, O-glycan biosynthesis, and alpha-synuclein signaling (Fig. 8C). These pathways above, therefore, are anticipated to be involved in the progression of LC. After that, the GO terms were also considered and determined for the corresponding genes.

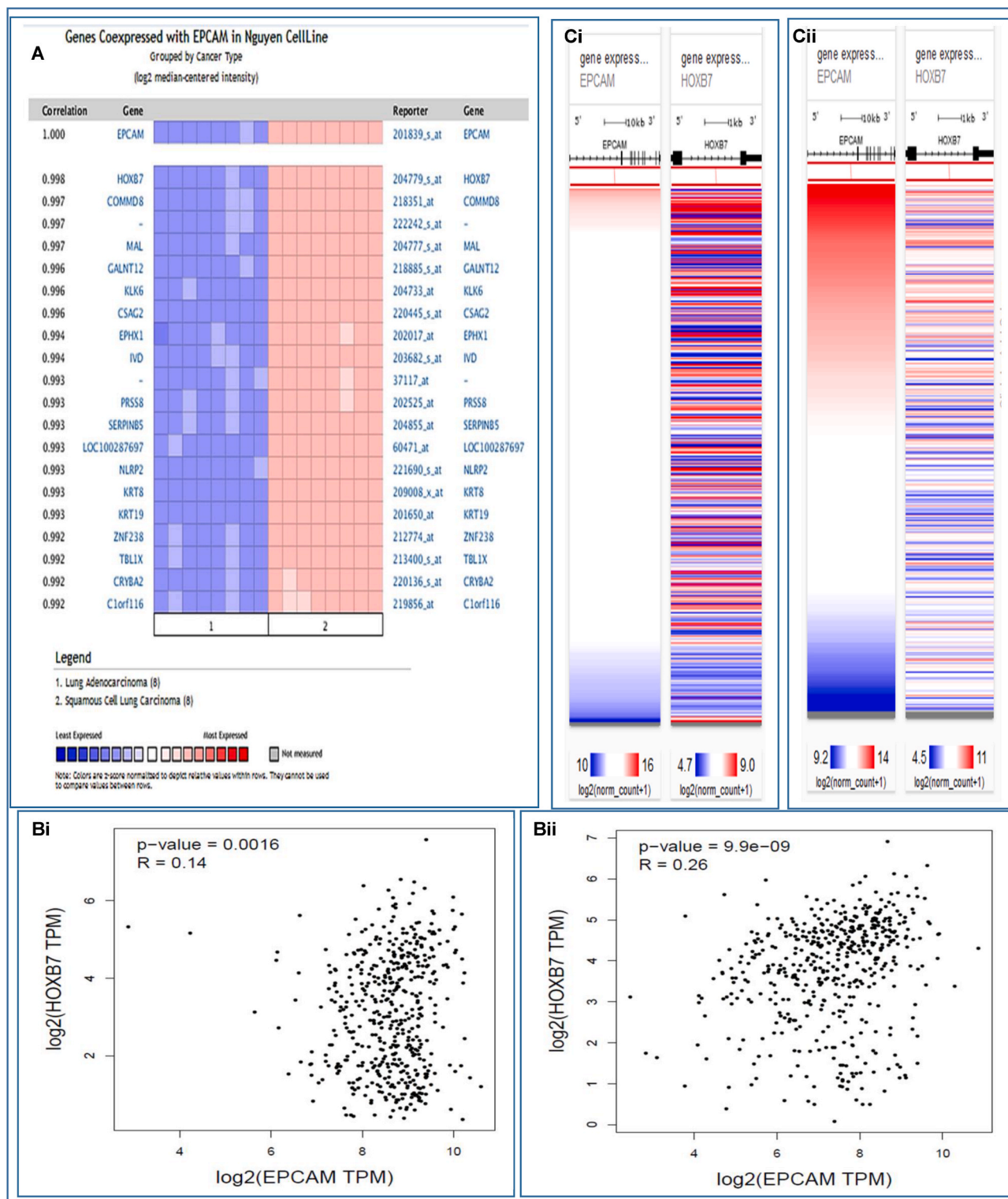


Fig. 7. Co-expression profile of the *EpCAM* and co-expressed genes in human LC. The figure shows (A) co-expression profile of *EpCAM* derived from the Oncomine database, (B) correlation analysis between *EpCAM* and *HOXB7* obtained by GEPIA2 server, (C) heatmap of mRNA expression for *EpCAM* and *HOXB7* genes across LC in the TCGA database, and (D) co-expression analysis between *EpCAM* and *HOXB7* genes in LC using UCSC Xena server.

The suggested GO features in the analysis mainly include myelination, membrane raft polarization and distribution, regulation of cell-cell adhesion, protein autoprocessing, regulation of ion transport, protein and fat metabolic process (Fig. 8D), hydrolase, dehydrogenase, peptidase and transferase (Fig. 8E), and mitotic spindle activity (Fig. 8F). The

cell signaling pathway where *EpCAM* gene is involved is depicted in Fig. 8G. These results suggested a possible correlation of *EpCAM* and its correlated genes in the development and progression of LC.

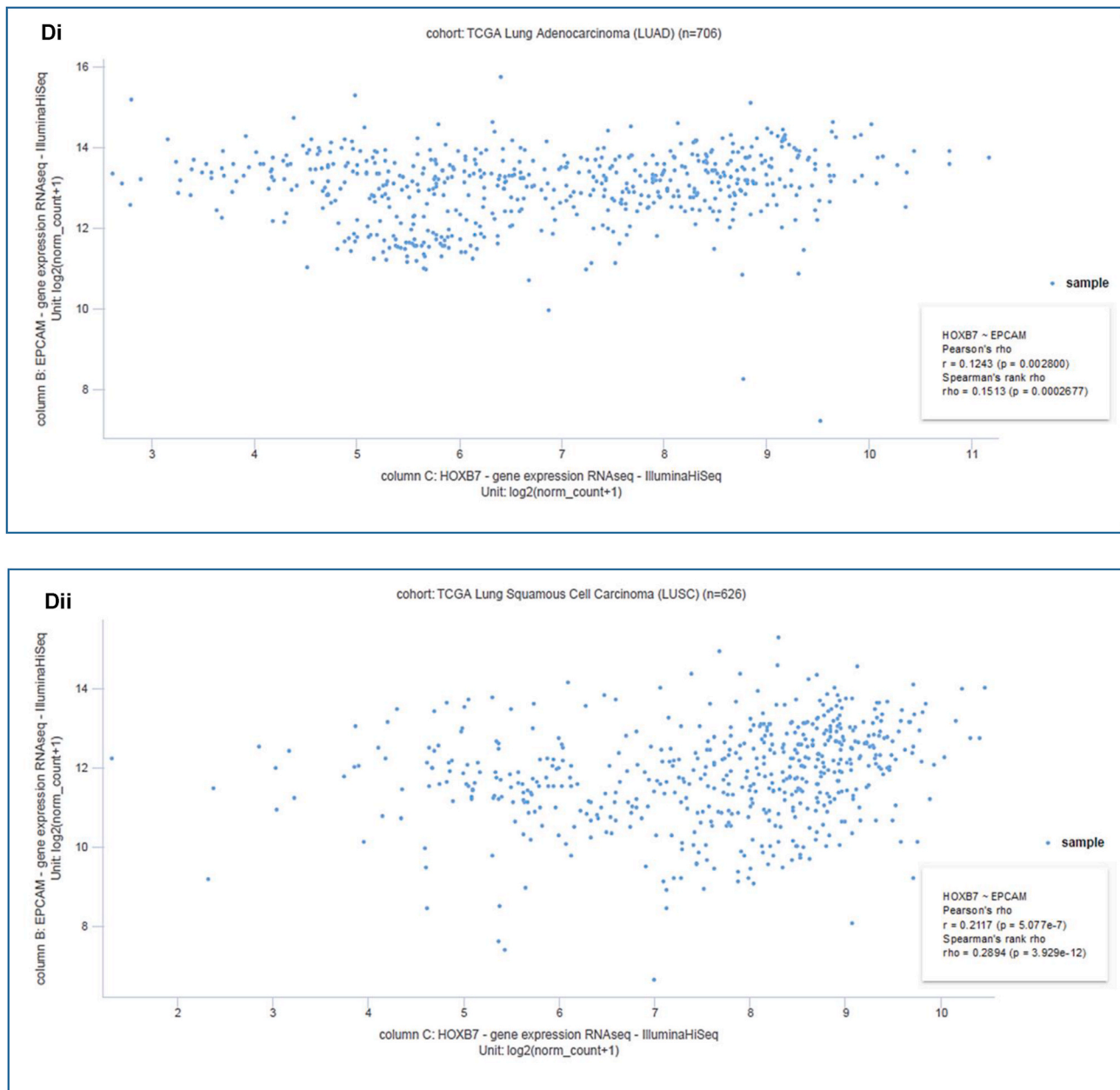


Fig. 7. (continued).

4. Discussion

LC has the highest cancer-related mortality rate worldwide and is infamously known for its appearance as one of the most frequently diagnosed malignant tumors [49,50]. NSCLC, consisting of two subtypes i.e., LUAD and LUSC, alone constitutes of an estimated 85% of all LC cases [50]. Hence, early detection and correct prognosis of LC are pivotal to understand a patient's illness circumstances and set up further appropriate treatment approaches, which is why the accurate prediction is crucial to both patients and physicians and researchers [51,52].

According to previous reports, LC in the majority of the patients (approx. 75%) seem to already reach the advanced stage (stage III/IV) during diagnosis which only puts a greater emphasis on the gravity of early diagnosis in case of LC [53]. For instance, surgical resection of NSCLC presents a favorable prognosis in case of small, localized tumors (stage I) with 5-year survival rates of 70–90% [54–56]. This suggests a strong correlation between the early diagnosis of LC and enhanced survival rates. Therefore, the early diagnosis of LC still remains a major

deciding factor for optimal outcomes, even with the addition of contemporary advancements and breakthroughs [57]. In this study, the significance of *EpCAM* as a prognostic marker in the early determination of two NSCLC i.e., LUAD and LUSC using bioinformatics was evaluated [33]. Based on the multiple databases, it can be concluded that the expression level of *EpCAM* is positively correlated with the progression of LC. The analysis of *EpCAM* expression in LUAD and LUSC displayed a negative correlation with all cases of overall survival, diseases free survival and relapse-free survival with the overall HR > 1. It was also observed that high *EpCAM* level showed a detrimental effect on the survival rate than those with lower *EpCAM* levels. A previous study suggested poor prognosis in LC patients due to increased expression of *EpCAM* which further validated this observation [33].

Again, the expression patterns of *EpCAM* in the two cancer tissues were also found to be significantly associated with different clinical characteristics of LC patients including, tumor histology, patient's race, gender, age, smoking habit, nodal metastasis status, etc. Therefore, analysis of these features requires further investigation since a high

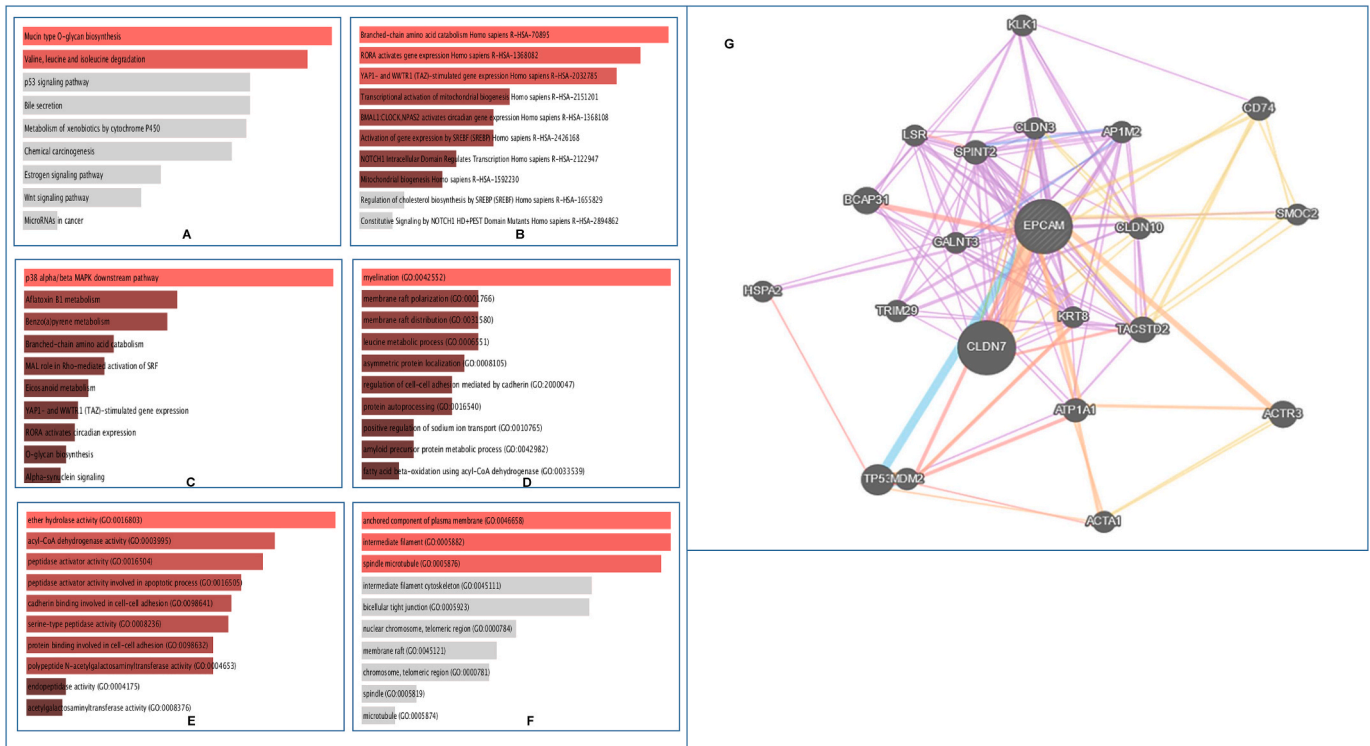


Fig. 8. Analysis of the pathways and gene ontologies related to *EpCAM* expression and LC. Pathways and ontologies are achieved from (A) KEGG human 2019, (B) Reactome 2016, (C) BioPlanet 2019, (D) GO biological process 2018, (E) GO molecular function 2018, and (F) GO cellular component 2018. The length and the color gradient of the bar represent the level of significance; brighter color indicates the more significant term and vice versa. (G) Figure representing the cell signaling network where *EpCAM* is involved. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

expression level of *EpCAM* might suggest a risk of cancer transformation and progression. Computer-aided systems and digital imaging, combined with the methodological aspects of modern immunohistochemistry, provide excellent insights into immunohistochemical scoring [33, 58,59]. The immunohistochemical data of *EpCAM* exhibited a strong nuclear immunoreactivity of *EpCAM* in every target LC cells. Robust and intense staining of cancer cells that distinguish them from normal alveolar cells, helped to recognize the higher levels of *EpCAM* expression in LUAD and LUSC tissues. Varied alterations of four categories i.e., somatically acquired genetic, epigenetic, transcriptomic, and proteomic alterations, constitute a series of histopathological process that contribute to cancer progression [58]. Any of these alterations in the genomic region (either loss or gain) can result in either suppressive or oncogenic effects [60]. The cBioPortal webserver was utilized to explore the copy number alterations, mutations, and mutant mRNA expressions of *EpCAM*. The *EpCAM* gene was found to be altered in 28 (<1%) of quarried LUAD samples with 0.5% somatic mutation frequency, whereas alteration was reported in 42 (4%) of the LUSC samples with 0.3% somatic mutation frequency. The highest mutation in both LUAD and LUSC was found to be a missense mutation, occurring in total of 1765 and 977 samples, respectively.

This study also investigated the potential correlation between DNA methylation and *EpCAM* expression which revealed a negative correlation between them. Transcriptional silencing of genes usually occurs as an outcome of DNA methylation that are found mainly on the CpG islands of the promoters of the genes. Following this prospect, several studies on the methylation of promoter in multiple cancer types including LC were conducted, which implied that the expression of *EpCAM* is inversely correlated with DNA methylation in tissues from cancer patients [61]. Through further co-expression and correlation analysis, it was found that 19 genes generated a positive correlation with the *EpCAM* gene and the homeobox B7 (*HOXB7*) was found to be positively correlated with *EpCAM* expression ($R = 0.998$). Analysis

conducted by other servers also confirmed the relationship between *EpCAM* and *HOXB7*. A dysregulated *HOXB7* was suggested to act as a critical player in regulating tumorigenesis and metastases of some cancers. A relevant previous study found the increased expression of *HOXB7* to be associated with poor clinical outcomes in LUAD patients leading to significant correlation with short survival time [62].

Finally, the correlated genes were used to analyze the possible *EpCAM* related pathways responsible for the development of LC. In the KEGG pathway analysis, the correlated genes were mostly related to the O-glycan biosynthesis and BCAA metabolism [63]. The observation seems reasonable as sometimes the process of oncogenesis is dependent on cellular energy and metabolites provided by amino acids that are degraded by overexpressed enzymes in many cancer types [64,65]. In the GO analysis, the widely used ontology term was myelination. Cognitive impairment is a common consequence of chemotherapy where myelination is reportedly claimed as an underlying factor according to previous studies [66]. Furthermore, membrane raft polarization facilitates another significant correlated molecular function which includes cell adhesion and migration. The cell adhesion disorders and aggressive phenotypes of migration and invasion constitute the malignant phenotype of cancer which is modulated by the dynamic feature of the cancer cell surface. Lipid rafts are also recently found to contribute to the cancer cell adhesion and migration [67]. Overall, the pathways and GO enrichment analysis portrayed the significance of *EpCAM* and its correlated genes in different oncogenic processes leading to LC development.

Overall, this study supports the idea that the expression of *EpCAM* gene could be used as the prognostic biomarker for the early detection of human LUAD and LUSC cases. However, more *in vivo* and *in vitro* studies are needed to finally confirm the outcome of this study.

5. Conclusion

In this study, we targeted the molecular signatures that play key roles in the development and progression of LUAD and LUSC. In cancer, prognostic factors are important for early diagnosis as well as efficient treatment and thus help patients prevent the risk of overtreatments. To determine the potency of *EpCAM* as a potential prognostic marker in LC development, the mRNA expression, DNA methylation, mutations and CNAs, correlated genes and the prognostic features were analyzed in this study. The analyses exhibited a sharp overexpression as well as a potential correlation of *EpCAM* to the LC development. Furthermore, it also indicates the probable signaling pathways and gene ontological features related to *EpCAM* and its expression in LC progression. These pathways could be the significant target to interfere with the development of cancer. For this reason, *EpCAM* is suggested to be an effective biomarker as well as a potential therapeutic target in effort of preventing LC in humans.

Data availability statement

Authors made all the data generated during experiment and analysis available within the manuscript.

Funding statement

Authors received no specific funding from any external sources.

Author contribution

BS and ATM conceived the study. BS designed the analysis. ATM, MAU and BS conducted the experiments. ATM, MAU and YA wrote the paper. NA, BS and BR edited the paper. BR supervised the study. BS was the co-supervisor of the study.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors are thankful to Swift Integrity Computational Lab, Dhaka, Bangladesh, for providing support during the preparation of the manuscript.

References

- [1] K.J. Harrington, *Biology of cancer*, *Medicine* 39 (12) (2011 Dec 1) 689–692, <https://doi.org/10.1016/j.mpmed.2011.09.015>.
- [2] Globocan, New global cancer data, [Internet]. Uicc.org. 2021 [cited 27 January 2021]. UICC (2020). Available from: <https://www.uicc.org/news/globocan-2020-new-global-cancer-data>.
- [3] C. Liu, X. Li, H. Shao, D. Li, Identification and validation of two lung adenocarcinoma-development characteristic gene sets for diagnosing lung adenocarcinoma and predicting prognosis, *Front. Genet.* 11 (2020), <https://doi.org/10.3389/fgene.2020.565206>. PMID:33408736 PMID:PMC7779611.
- [4] M.C. Wong, X.Q. Lao, K.F. Ho, W.B. Goggins, L.A. Shelly, Incidence and mortality of lung cancer: global trends and association with socioeconomic status, *Sci. Rep.* 7 (1) (2017) 1–9, <https://doi.org/10.1038/s41598-017-02997-2>. PMID:28361988 PMID:PMC5374459, 10.1038/s41598-017-02997-2, 10.1038/srep45846.
- [5] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA A Cancer J. Clin.* 68 (6) (2018) 394–424, <https://doi.org/10.3322/caac.21492>. PMID:30207593.
- [6] D. Anusewicz, M. Orzechowska, A.K. Bednarek, Lung squamous cell carcinoma and lung adenocarcinoma differential gene expression regulation through pathways of Notch, Hedgehog, Wnt, and ErbB signalling, *Sci. Rep.* 10 (1) (2020) 1–15, <https://doi.org/10.1038/s41598-020-77284-8>. PMID:33273537 PMID:PMC7713208.
- [7] Cancer Genome Atlas Research Network, Comprehensive genomic characterization of squamous cell lung cancers, *Nature* 489 (7417) (2012) 519, <https://doi.org/10.1038/nature11404>. PMID:22960745 PMID:PMC3466113.
- [8] I. Donner, R. Katainen, L.J. Sipilä, M. Aavikko, E. Pukkala, L.A. Aaltonen, Germline mutations in young non-smoking women with lung adenocarcinoma, *Lung Canc.* 122 (2018) 76–82, <https://doi.org/10.1016/j.lungcan.2018.05.027>. PMID:30032850.
- [9] D.J. Ling, Z.S. Chen, Q.D. Liao, J.X. Feng, X.Y. Zhang, T.Y. Yin, Differential effects of MTTSS1 on invasion and proliferation in subtypes of non-small cell lung cancer cells, *Experimental and therapeutic medicine* 12 (2) (2016) 1225–1231, <https://doi.org/10.3892/etm.2016.3382>. PMID:27446348 PMID:PMC4950550.
- [10] T. Okegawa, R.C. Pong, Y. Li, J.T. Hsieh, The role of cell adhesion molecule in cancer progression and its application in cancer therapy, *Acta Biochim. Pol.* 51 (2) (2004) 445–457, <https://doi.org/10.18388/abp.2004.3583>. PMID:15218541.
- [11] G.M. Calaf, D. Roy, G. Narayan, A.S. Balajee, Differential expression of cell adhesion molecules in an ionizing radiation-induced breast cancer model system, *Oncol. Rep.* 30 (1) (2013) 285–291, <https://doi.org/10.3892/or.2013.2448>. PMID:23670055.
- [12] X.W. Xi, J.L. Chen, Y.J. Feng, X.P. Wan, K.J. Gu, Differential expression of adhesion molecules (CD44, CD29, ICAM-1 and E-cadherin) in ovarian cancer SK-OV-3ip1 cells grown as monolayer and multicellular aggregates, *Chin. J. Canc. Res.* 15 (1) (2003) 19–23, <https://doi.org/10.1007/s11670-003-0005-3>.
- [13] M. Bujko, P. Kober, M. Mikula, M. Ligaj, J. Ostrowski, J.A. Siedlecki, Expression changes of cell adhesion related genes in colorectal tumors, *Oncology letters* 9 (6) (2015) 2463–2470, <https://doi.org/10.3892/ol.2015.3107>. PMID:26137091 PMID:PMC4473523.
- [14] M.C. Boelens, A. van den Berg, I. Vogelzang, J. Wesseling, D.S. Postma, W. Timens, H.J. Groen, Differential expression and distribution of epithelial adhesion molecules in non-small cell lung cancer and normal bronchus, *J. Clin. Pathol.* 60 (6) (2007) 608–614, <https://doi.org/10.1136/jcp.2005.031443>. PMID:16489176 PMID:PMC1955047.
- [15] C. Schardt, J. Heymanns, C. Schardt, M. Rotsch, K. Havemann, Differential expression of the intercellular adhesion molecule-1 (ICAM-1) in lung cancer cell lines of various histological types, *Eur. J. Canc.* 29 (16) (1993) 2250–2255, [https://doi.org/10.1016/0959-8049\(93\)90217-4](https://doi.org/10.1016/0959-8049(93)90217-4).
- [16] T. Hase, M. Sato, K. Yoshida, L. Girard, Y. Takeyama, M. Horio, Y. Hasegawa, Pivotal role of epithelial cell adhesion molecule in the survival of lung cancer cells, *Canc. Sci.* 102 (8) (2011) 1493–1500, <https://doi.org/10.1111/j.1349-7006.2011.01973.x>. PMID:21535318 PMID:PMC3381954.
- [17] Y. Yu, X. Tian, Analysis of genes associated with prognosis of lung adenocarcinoma based on GEO and TCGA databases, *Medicine* 99 (19) (2020), <https://doi.org/10.1097/MD.00000000000020183>. PMID:32384511 PMID:PMC7220259.
- [18] B. Thirion, S. Pereira, A. Neveol, B. Dahamna, S. Darmoni, October). French MeSH Browser: a cross-language tool to access MEDLINE/PubMed, in: *AMIA... Annual Symposium Proceedings. AMIA Symposium, 2007*, 1132–1132.
- [19] M. Janiszewska, M.C. Primi, T. Izard, Cell adhesion in cancer: beyond the migration of single cells, *J. Biol. Chem.* 295 (8) (2020) 2495–2505, <https://doi.org/10.1074/jbc.REV119.007759>. PMID:31937589 PMID:PMC7039572.
- [20] H. Lodish, A. Berk, S.L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, *Cell-cell adhesion and communication*, in: *Molecular Cell Biology*, fourth ed., WH Freeman, 2000.
- [21] L. Huang, Y. Yang, F. Yang, S. Liu, Z. Zhu, Z. Lei, J. Guo, Functions of *EpCAM* in physiological processes and diseases, *Int. J. Mol. Med.* 42 (4) (2018) 1771–1785, <https://doi.org/10.3892/ijmm.2018.3764>. PMID:30015855 PMID:PMC6108866.
- [22] M. Herlyn, Z. Steplewski, D. Herlyn, H. Koprowski, Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies, *Proc. Natl. Acad. Sci. Unit. States Am.* 76 (3) (1979) 1438–1442, <https://doi.org/10.1073/pnas.76.3.1438>. PMID:286328 PMID:PMC383267.
- [23] U. Schnell, V. Cirulli, B.N. Giepmans, *EpCAM*: structure and function in health and disease, *Biochim. Biophys. Acta Biomembr.* 1828 (8) (2013) 1989–2001, <https://doi.org/10.1016/j.bbmem.2013.04.018>. PMID:23618806.
- [24] K. Kamimoto, K. Kaneko, C.Y.Y. Kok, H. Okada, A. Miyajima, T. Itoh, Heterogeneity and stochastic growth regulation of biliary epithelial cells dictate dynamic epithelial tissue remodeling, *Elife* 5 (2016), e15034, <https://doi.org/10.7554/eLife.15034.024>.
- [25] E. Schmelzer, L. Zhang, A. Bruce, E. Wauthier, J. Ludlow, H.L. Yao, L.M. Reid, Human hepatic stem cells from fetal and postnatal donors, *J. Exp. Med.* 204 (8) (2007) 1973–1987, <https://doi.org/10.1084/jem.20061603>. PMID:17664288 PMID:PMC2118675.
- [26] M. Trzpis, P.M. McLaughlin, L.M. de Leij, M.C. Harmsen, Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule, *Am. J. Pathol.* 171 (2) (2007) 386–395, <https://doi.org/10.2353/ajpath.2007.070152>. PMID:17600130 PMID:PMC1934518.
- [27] W.W. Agace, J.M. Higgins, B. Sadasivan, M.B. Brenner, C.M. Parker, T-lymphocyte-epithelial-cell interactions: integrin αE (CD103) $\beta 7$, LEEP-CAM and chemokines, *Curr. Opin. Cell Biol.* 12 (5) (2000) 563–568, [https://doi.org/10.1016/S0955-0674\(00\)00132-0](https://doi.org/10.1016/S0955-0674(00)00132-0).
- [28] G. Moldenhauer, A.V. Salnikov, S. Lüttgau, I. Herr, J. Anderl, H. Faulstich, Therapeutic potential of amantitin-conjugated anti-epithelial cell adhesion molecule monoclonal antibody against pancreatic carcinoma, *J. Natl. Cancer Inst.* 104 (8) (2012) 622–634, <https://doi.org/10.1093/jnci/djs140>. PMID:22457476.
- [29] A. Armstrong, S.L. Eck, *EpCAM*: a new therapeutic target for an old cancer antigen, *Canc. Biol. Ther.* 2 (4) (2003) 320–326, <https://doi.org/10.4161/cbt.2.4.451>. PMID:14508099.
- [30] S. Han, S. Zong, Q. Shi, H. Li, S. Liu, W. Yang, F. Hou, Is *Ep-CAM* expression a diagnostic and prognostic biomarker for colorectal cancer? A systematic meta-analysis, *EBioMedicine* 20 (2017) 61–69, <https://doi.org/10.1016/j.ebiom.2017.05.025>. PMID:28558958 PMID:PMC5478257.
- [31] G. Spizzo, G. Gastl, P. Obrist, P. Went, S. Dirnhofer, S. Bischoff, H. Dietrich, High *Ep-CAM* expression is associated with poor prognosis in node-positive breast

- cancer, *Breast Canc. Res. Treat.* 86 (3) (2004) 207–213, <https://doi.org/10.1023/B:BREA.0000036787.59816.01>. PMID:15567937.
- [32] G. Spizzo, P. Went, S. Dirnhofer, P. Obrist, H. Moch, P.A. Baeuerle, A.G. Zeimet, Overexpression of epithelial cell adhesion molecule (Ep-CAM) is an independent prognostic marker for reduced survival of patients with epithelial ovarian cancer, *Gynecol. Oncol.* 103 (2) (2006) 483–488, <https://doi.org/10.1016/j.ygyno.2006.03.035>. PMID:16678891.
- [33] M.G. Pak, D.H. Shin, C.H. Lee, M.K. Lee, Significance of *EpCAM* and *TROP2* expression in non-small cell lung cancer, *World J. Surg. Oncol.* 10 (2012 Apr 6) 53, <https://doi.org/10.1186/1477-7819-10-53>. PMID:22482828 PMCid:PMC3342122.
- [34] D.B. Seligson, A.J. Pantuck, X. Liu, Y. Huang, S. Horvath, M.H. Bui, R.A. Figlin, Epithelial cell adhesion molecule (KSA) expression: pathobiology and its role as an independent predictor of survival in renal cell carcinoma, *Clin. Canc. Res.* 10 (8) (2004) 2659–2669, <https://doi.org/10.1158/1078-0432.CCR-1132-03>. PMID:15102668.
- [35] S. Crvenkova, Prognostic factors and survival in NSCLC patients treated with chemoradiotherapy, *Open access Macedonian journal of medical sciences* 3 (1) (2015) 75, <https://doi.org/10.3889/oamjms.2015.003>. PMID:27275200 PMCid:PMC4877793.
- [36] H. Lemjabbar-Alaoui, O.U. Hassan, Y.W. Yang, P. Buchanan, Lung cancer: biology and treatment options, *Biochim. Biophys. Acta Rev. Canc* 1856 (2) (2015) 189–210, <https://doi.org/10.1016/j.bbcan.2015.08.002>. PMID:26297204 PMCid:PMC4663145.
- [37] Y. Ma, M. Fan, L. Dai, X. Kang, Y. Liu, Y. Sun, K. Chen, The expression of TTF-1 and Napsin A in early-stage lung adenocarcinoma correlates with the results of surgical treatment, *Tumor Biol.* 36 (10) (2015) 8085–8092, <https://doi.org/10.1007/s13277-015-3478-z>, PMID:25982999.
- [38] D.R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, A.M. Chinnaiyan, ONCOMINE: a cancer microarray database and integrated data-mining platform, *Neoplasia* 6 (1) (2004) 1, [https://doi.org/10.1016/S1476-5586\(04\)80047-2](https://doi.org/10.1016/S1476-5586(04)80047-2).
- [39] Z. Tang, B. Kang, C. Li, T. Chen, Z. Zhang, GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis, *Jul 2;47(W1):W556-60, Nucleic Acids Res.* (2019), <https://doi.org/10.1093/nar/gkz430>. PMID:31114875 PMCid:PMC6602440.
- [40] S.J. Park, B.H. Yoon, S.K. Kim, S.Y. Kim, GENT2: an updated gene expression database for normal and tumor tissues, *BMC Med. Genom.* 12 (5) (2019 Jul) 101, <https://doi.org/10.1186/s12920-019-0514-7>. PMID:31296229 PMCid:PMC6624177.
- [41] D.S. Chandrashekar, B. Bashel, S.A. Balasubramanya, C.J. Creighton, I. Ponce-Rodriguez, B.V. Chakravarthi, S. Varambally, UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, *Neoplasia* 19 (8) (2017 Aug 1) 649–658, <https://doi.org/10.1016/j.neo.2017.05.002>, PMID:28732212 PMCid:PMC5516091.
- [42] M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, Tissue-based map of the human proteome, *Science* (6220) (2015 Jan 23) 347, <https://doi.org/10.1126/science.1260419>. PMID:25613900.
- [43] M.J. Goldman, B. Craft, M. Hastie, K. Repecka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A.N. Brooks, J. Zhu, Visualizing and interpreting cancer genomics data via the Xena platform, *Nat. Biotechnol.* (2020 May 22) 1–4, <https://doi.org/10.1038/s41587-020-0546-8>. PMID:32444850 PMCid:PMC7386072.
- [44] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C. J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, The cBio Cancer Genomics Portal: an Open Platform for Exploring Multidimensional Cancer Genomics Data, 2012, <https://doi.org/10.1158/2159-8290.CD-12-0095>. PMID:22588877 PMCid:PMC3956037.
- [45] H. Mizuno, K. Kitada, K. Nakai, A. Sarai, PrognoScan: a new database for meta-analysis of the prognostic value of genes, *BMC Med. Genom.* 2 (1) (2009 Dec 1) 18, <https://doi.org/10.1186/1755-8794-2-18>, PMID:19393097 PMCid:PMC2689870.
- [46] E.Y. Chen, C.M. Tan, Y. Kou, Q. Duan, Z. Wang, G.V. Meirelles, N.R. Clark, A. Ma'ayan, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, *BMC Bioinf.* 14 (1) (2013 Dec 1) 128, <https://doi.org/10.1186/1471-2105-14-128>. PMID:23586463 PMCid:PMC3637064.
- [47] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, W90-7, *Nucleic Acids Res.* 44 (W1) (2016 Jul 8), <https://doi.org/10.1093/nar/gkw377>. PMID:27141961 PMCid:PMC4987924.
- [48] T. Phillips, The role of methylation in gene expression, *Nature Education* 1 (1) (2008) 116.
- [49] C.S.D. Cruz, L.T. Tanoue, R.A. Matthay, Lung cancer: epidemiology, etiology, and prevention, *Clin. Chest Med.* 32 (4) (2011) 605–644, <https://doi.org/10.1016/j.ccm.2011.09.001>. PMID:22054876 PMCid:PMC3864624.
- [50] Y. Yang, M. Wang, B. Liu, Exploring and comparing of the gene expression and methylation differences between lung adenocarcinoma and squamous cell carcinoma, *J. Cell. Physiol.* 234 (4) (2019) 4454–4459, <https://doi.org/10.1002/jcp.27240>. PMID:30317601.
- [51] E.H. Bernicker, T.C. Allen, P.T. Cagle, Update on emerging biomarkers in lung cancer, *J. Thorac. Dis.* 11 (Suppl 1) (2019) S81, <https://doi.org/10.21037/jtd.2019.01.46>. PMID:30775031 PMCid:PMC6353743.
- [52] K.G.M. Moons, D.G. Altman, Y. Vergouwe, P. Royston, Prognosis and prognostic research: application and impact of prognostic models in clinical practice, *b606, b606, BMJ (Clinical research ed.)* 338 (2009), <https://doi.org/10.1136/bmj.b606>. PMID:19502216.
- [53] S. Walters, et al., Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004–2007, *Thorax* 68 (2013) 551–564, <https://doi.org/10.1136/thoraxjnl-2012-202297>. PMID:23399908.
- [54] P. Goldstraw, et al., The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer, *J. Thorac. Oncol.* 11 (2016) 39–51, <https://doi.org/10.1016/j.jtho.2015.09.009>.
- [55] J.C. Nesbitt, J.B. Putnam Jr., G.L. Walsh, J.A. Roth, C.F. Mountain, Survival in early-stage non-small cell lung cancer, *Ann. Thorac. Surg.* 60 (1995) 466–472, [https://doi.org/10.1016/0003-4975\(95\)00169-L](https://doi.org/10.1016/0003-4975(95)00169-L).
- [56] R. Shah, S. Sabanathan, J. Richardson, A.J. Mearns, C. Goulden, Results of surgical treatment of stage I and II lung cancer, *J. Cardiovasc. Surg.* 37 (1996) 169–172.
- [57] S.H. Bradley, M.P. Kennedy, R.D. Neal, Recognising lung cancer in primary care, *Adv. Ther.* 36 (1) (2019) 19–30, <https://doi.org/10.1007/s12325-018-0843-5>. PMID:30499068 PMCid:PMC6318240.
- [58] A. Laurinavicius, A. Laurinaviciene, V. Ostapenko, D. Dasevicius, S. Jarmalaite, J. Lazutka, Immunohistochemistry profiles of breast ductal carcinoma: factor analysis of digital image analysis data, 27, 27, *Diagn. Pathol.* 7 (2012), <https://doi.org/10.1186/1746-1596-7-27>. PMID:22424533 PMCid:PMC3319425.
- [59] L. Yu, Z. Xiao, H. Tu, B. Tong, S. Chen, The expression and prognostic significance of Drp1 in lung cancer: a bioinformatics analysis and immunohistochemistry, *e18228, e18228, Medicine* 98 (48) (2019), <https://doi.org/10.1097/MD.00000000000018228>. PMID:31770286 PMCid:PMC6890372.
- [60] K. Klonowska, K. Czubak, M. Wojciechowska, L. Handschuh, A. Zmienko, M. Figlerowicz, H. Dams-Kozłowska, P. Kozłowski, Oncogenomic portals for the visualization and analysis of genome-wide cancer data, *Oncotarget* 7 (1) (2016) 176–192, <https://doi.org/10.18632/oncotarget.6128>. PMID:26484415 PMCid:PMC4807991.
- [61] L. Keller, S. Werner, K. Pantel, Biology and clinical relevance of EpCAM, *Cell Stress* 3 (6) (2019) 165, <https://doi.org/10.15698/cst2019.06.188>. PMID:31225512 PMCid:PMC6558934.
- [62] W. Yuan, X. Zhang, Y. Xu, S. Li, Y. Hu, S. Wu, Role of HOXB7 in regulation of progression and metastasis of human lung adenocarcinoma, *Mol. Carcinog.* 53 (1) (2014) 49–57, <https://doi.org/10.1002/mc.21947>. PMID:22911672.
- [63] A. Vazquez, J.J. Kamphorst, E.K. Markert, et al., Cancer metabolism at a glance, *J. Cell Sci.* 129 (2016) 3367–3373, <https://doi.org/10.1242/jcs.181016>. PMID:27635066 PMCid:PMC6518336.
- [64] E. Ananieva, Targeting amino acid metabolism in cancer growth and antitumor immune response, *World J. Biol. Chem.* 6 (2015) 281–289, <https://doi.org/10.4331/wjbc.v6.i4.281>. PMID:26629311 PMCid:PMC4657121.
- [65] R.J. DeBerardinis, N.S. Chandel, Fundamentals of cancer metabolism [PMC free article] [PubMed] [Google Scholar]The recent review by DeBerardinis and Chandel provides a useful overview of the role of metabolism in cancer, *Sci Adv* 2 (2016), e1600200, <https://doi.org/10.1126/sciadv.1600200>. PMID:27386546 PMCid:PMC4928883.
- [66] T.L. Briones, J. Woods, Dysregulation in myelination mediated by persistent neuroinflammation: possible mechanisms in chemotherapy-related cognitive impairment, *Brain Behav. Immun.* 35 (2014) 23–32, <https://doi.org/10.1016/j.bbi.2013.07.175>. PMID:23916895 PMCid:PMC3858476.
- [67] T. Murai, The role of lipid rafts in cancer cell adhesion and migration, *International journal of cell biology* (2012), <https://doi.org/10.1155/2012/763283>. PMID:22253629 PMCid:PMC3255102.